CLAIMS

1. A method comprising:

separating at least a portion of an audio signal into a plurality of frames;

extracting line spectrum pairs from each of the plurality of frames; and

using at least the line spectrum pairs to classify at least the portion as either speech or non-speech.

2. A method as recited in claim 1, wherein the using comprises:

generating an input Gaussian Model corresponding to the plurality of frames based on the extracted line spectrum pairs;

comparing the input Gaussian Model to a Vector Quantization codebook including a plurality of trained Gaussian Models;

identifying one of the plurality of trained Gaussian Models that is closest to the input Gaussian Model;

determining a distance between the input Gaussian Model and the closest trained Gaussian Model; and

classifying at least the portion as speech if the distance is less than a threshold value.

3. A method as recited in claim 1, wherein the using comprises:

generating an input Gaussian Model corresponding to the plurality of frames based on the extracted line spectrum pairs;

identifying one of the plurality of trained Gaussian Models that is closest to the input Gaussian Model;

determining a distance between the input Gaussian Model and the closest trained Gaussian Model; and

classifying at least the portion as non-speech if the distance is greater than a first threshold value.

4. A method as recited in claim 3, further comprising:

determining an energy distribution of the plurality of frames in a first bandwidth; and

classifying at least the portion as non-speech if the distance is greater than a second threshold value and the energy distribution of the plurality of frames in the first bandwidth is less than a third threshold value, wherein the second threshold value is less than the first threshold value.

5. A method as recited in claim 4, further comprising:

determining an energy distribution of the plurality of frames in a second bandwidth; and

classifying at least the portion as speech if the distance is less than the second threshold value and the energy distribution of the plurality of frames in the second bandwidth is greater than a fourth threshold value.

6. A method as recited in claim 5, further comprising otherwise classifying at least the portion as speech.

7. A method as recited in claim 2, further comprising:

extracting a high zero crossing rate ratio feature from the plurality of frames;

extracting a low short time energy ratio feature from the plurality of frames;

extracting a spectrum flux feature from the plurality of frames;

pre-classifying the portion as speech or non-speech based at least in part on an average zero crossing rate, the high zero crossing rate ratio, the low short time energy ratio, and the spectrum flux features;

using a first value as the threshold value if the portion is pre-classified as speech; and

using a second value as the threshold value if the portion is pre-classified as non-speech, wherein the second value is greater than the first value.

8. One or more computer-readable memories containing a computer program that is executable by a processor to perform the method recited in claim 1.

9. A method comprising:

separating at least a portion of an audio signal into a plurality of frames;

extracting a periodicity feature from the plurality of frames; and

using at least the periodicity feature to classify at least the portion as either music or environment sound.

10. A method as recited in claim 9, wherein the periodicity feature comprises a noise frame ratio that identifies a ratio of noise frames to non-noise frames in the plurality of frames.

11. A method as recited in claim 10, further comprising classifying at least the portion as environment sound if the noise frame ratio exceeds a threshold value.

12. A method as recited in claim 10, further comprising:

extracting, from the plurality of frames, a band periodicity for each of a plurality of bands of the audio signal and a full band periodicity that is a concatenation of the band periodicities for each of the plurality of bands; and

classifying at least the portion as environment sound if the first band periodicity is less than a first threshold or the second band is less than a second threshold.

13. A method as recited in claim 9, wherein the periodicity feature comprises a band periodicity for each of a plurality of bands of the audio signal.

14. A method as recited in claim 13, further comprising:

extracting a full band periodicity from the plurality of frames that is a concatenation of the band periodicities for each of the plurality of bands; and

classifying at least the portion as environment sound if the full band periodicity exceeds a threshold value.

15. A method as recited in claim 9, further comprising extracting a spectrum flux feature from the plurality of frames, and wherein the using comprises using at least the periodicity feature and the spectrum flux feature to classify at least the portion as either music or environment sound.

16. A method as recited in claim 15, wherein the spectrum flux feature is extracted by determining a Fast Fourier Transform for each of the plurality of frames and calculating a difference in the Fast Fourier Transforms for successive frames.

17. A method as recited in claim 15, further comprising:

extracting, from the plurality of frames, an energy distribution in a band of the audio signal; and

classifying at least the portion as environment sound if the band energy distribution is less than a first threshold or the spectrum flux exceeds a second threshold.

18. A method as recited in claim 9, wherein the periodicity feature comprises a band periodicity for each of a plurality of bands of the audio signal, and further comprising:

extracting, from the plurality of frames, a spectrum flux feature;

extracting, from the plurality of frames, an energy feature indicating an amount of energy in at least one band of the portion; and

classifying at least the portion as environment sound if the amount of energy is less than a first threshold and the spectrum flux is less than a second threshold.

19. One or more computer-readable memories containing a computer program that is executable by a processor to perform the method recited in claim 9.

20. A method comprising:

separating at least a portion of an audio signal into a plurality of frames;

extracting a periodicity feature for each of the plurality of frames; and

using at least the periodicity feature to classify the plurality of frames as either music with vocals or music without vocals.

21. A method as recited in claim 20, wherein the periodicity feature comprises a band periodicity for each of a plurality of bands of the audio signal.

22. A method as recited in claim 21, further comprising classifying at least the portion as music with vocals if the band periodicity of at least one of the plurality of bands is greater than a first threshold and less than a second threshold.

23. A method as recited in claim 22, further comprising classifying at least the portion as environment sound if the band periodicity of each of the plurality of bands is less than the second threshold, and otherwise classifying at least the portion as music without vocals.

24. One or more computer-readable memories containing a computer program that is executable by a processor to perform the method recited in claim 20.

25. A method for determining when a speaker changes, the method comprising:

separating at least a portion of an audio signal into a plurality of frames;

extracting line spectrum pairs from each of the plurality of frames; and

determining when a speaker of the audio signal changes based at least in part on the line spectrum pairs.

26. A method as recited in claim 25, wherein the determining comprises:

calculating a difference between line spectrum pairs for successive frames of the plurality of frames;

if the difference between two line spectrum pairs exceeds a threshold value, then determining that the speaker has changed, otherwise determining that the speaker has not changed.

27. One or more computer-readable memories containing a computer program that is executable by a processor to perform the method recited in claim 25.

28. An apparatus comprising:

a line spectrum pair (LSP) analyzer to extract line spectrum pairs from a portion of an audio signal; and

a speech discriminator, communicatively coupled to the LSP analyzer, to classify the portion of the audio signal as either speech or non-speech based at least in part on the LSP analyzer.

29. An apparatus as recited in claim 28, further comprising:

a distance calculator, communicatively coupled to the LSP analyzer, to determine a distance between at least one of the trained Gaussian Models and an input Gaussian Model based on the extracted line spectrum pairs; and

wherein the speech discriminator is further to classify the portion of the audio signal as either speech or non-speech based at least in part on the distance between the at least one of the trained Gaussian Models and the input Gaussian Model.

30. An apparatus as recited in claim 28, further comprising:

a Fast Fourier Transform (FFT) analyzer to extract Fast Fourier Transform features from the portion of the audio signal;

an energy distribution calculator, communicatively coupled to both the FFT analyzer and the speech discriminator, to determine an energy distribution of the portion of the audio signal in at least one bandwidth; and

wherein the speech discriminator is further to classify the portion of the audio signal as either speech or non-speech based at least in part on the energy distribution of the portion of the audio signal in the at least one bandwidth.

31. An apparatus comprising:

a band periodicity calculator to determine a periodicity of each of a plurality of bands of a portion of an audio signal; and

a discriminator, communicatively coupled to the band periodicity calculator, to classify the portion of the audio signal as music or environment sound based at least in part on the periodicity of one of the plurality of bands.

32. An apparatus as recited in claim 31, further comprising:

a noise frame ratio calculator, communicatively coupled to the discriminator, to determine a noise frame ratio of the portion of the audio signal; and

wherein the discriminator is to classify the portion of the audio signal as music or environment sound based at least in part on the periodicity of one of the plurality of bands and on the noise frame ratio of the portion.

33. An apparatus as recited in claim 31, further comprising:

a spectrum flux analyzer, communicatively coupled to the discriminator, to determine a spectrum flux of the portion of the audio signal; and

wherein the discriminator is to classify the portion of the audio signal as music or environment sound based at least in part on the periodicity of one of the plurality of bands and on the spectrum flux of the portion.

34. A method comprising:

receiving an audio signal;

separating the audio signal into a plurality of portions; and

classifying each of the plurality of portions, based at least in part on periodicity features of the portion, as one of: speech, music, silence, and environment sound.

35. A method as recited in claim 34, wherein the periodicity features include a noise frame ratio that identifies a ratio of noise frames to non-noise frames in the plurality of frames.

36. A method as recited in claim 35, wherein the classifying comprises classifying at least the portion as environment sound if the noise frame ratio exceeds a threshold value.

37. A method as recited in claim 34, further comprising:

extracting, from the plurality of frames, a band periodicity for each of a plurality of bands of the audio signal and a full band periodicity that is a concatenation of the band periodicities for each of the plurality of bands; and

wherein the classifying comprises classifying at least the portion as environment sound if a band periodicity of a first of the plurality of bands is less

than the first threshold a band periodicity of a second of the plurality of bands is less than the second threshold. .

38. A method as recited in claim 34, wherein the periodicity features include a band periodicity for each of a plurality of bands of the audio signal.

39. A method as recited in claim 38, further comprising:

extracting a full band periodicity from the plurality of frames that is a concatenation of the band periodicities for each of the plurality of bands; and

wherein the classifying comprises classifying at least the portion as environment sound if the full band periodicity exceeds a threshold value.

40. A method as recited in claim 34, further comprising:

extracting a spectrum flux feature from the plurality of frames; and

wherein the classifying comprises classifying at least the portion as either music or environment sound based at least in part on the periodicity feature and the spectrum flux feature.

41. A method as recited in claim 34, further comprising:

extracting line spectrum pairs from each of the plurality of frames;

generating an input Gaussian Model corresponding to the plurality of frames based on the extracted line spectrum pairs;

comparing the input Gaussian Model to a Vector Quantization codebook including a plurality of trained Gaussian Models;

identifying one of the plurality of trained Gaussian Models that is closest to the input Gaussian Model;

determining a distance between the input Gaussian Model and the closest trained Gaussian Model; and

classifying at least the portion as speech if the distance is less than a threshold value.

42. A method as recited in claim 34, further comprising:

extracting line spectrum pairs from each of the plurality of frames;

generating an input Gaussian Model corresponding to the plurality of frames based on the extracted line spectrum pairs;

identifying one of the plurality of trained Gaussian Models that is closest to the input Gaussian Model;

determining a distance between the input Gaussian Model and the closest trained Gaussian Model; and

classifying at least the portion as one of music, silence, or environment sound if the distance is greater than a first threshold value.

43. A method as recited in claim 42, further comprising:

determining an energy distribution of the plurality of frames in a first bandwidth; and

classifying at least the portion as one of music, silence, or environment sound if the distance is greater then a second threshold value and the energy distribution of the plurality of frames in the first bandwidth is less than a third

threshold value, wherein the second threshold value is less than the first threshold value.

44. A method as recited in claim 43, further comprising:

determining an energy distribution of the plurality of frames in a second bandwidth; and

classifying at least the portion as one of music, silence, or environment sound if the distance is greater than a fourth threshold value and the energy distribution of the plurality of frames in the second bandwidth is less than a fifth threshold value, wherein the fourth threshold value is less than the first threshold value.

45. A method as recited in claim 44, further comprising otherwise classifying at least the portion as speech.

46. One or more computer-readable memories containing a computer program that is executable by a processor to perform the method recited in claim 34.

47. One or more computer-readable media having stored thereon a computer program to classify a portion of an audio signal as speech, music, silence, or environment sound, wherein the computer program, when executed by one or more processors, causes the one or more processors to perform acts including:

(a) analyzing line spectrum pair features of the portion to determine if the portion is speech;

(b) analyzing energy features of the portion to determine if the portion is silence;

(c) analyzing periodicity features of the portion to determine if the portion is music or environment sound; and

(d) classifying the portion as speech, music, silence, or environment sound based on at least one of the analyzing acts (a)-(c).

48. One or more computer-readable media as recited in claim 47, wherein the computer program is further to cause the one or more processors to perform the acts (a) – (d) in the order (a), then (b), then (c), then (d).

49. One or more computer-readable media as recited in claim 48, wherein the computer program is further to cause the one or more processors to perform act (b) only if act (a) results in a determination that the portion is not speech.

50. One or more computer-readable media as recited in claim 48, wherein the computer program is further to cause the one or more processors to perform act (c) only if act (b) results in a determination that the portion is not silence.